Regulating ChatGPT and other Large Generative AI Models

Year: 2023 | *Citations: 403* | *Authors: P. Hacker, A. Engel, M. Mauer*

Abstract

Large generative AI models (LGAIMs), such as ChatGPT, GPT-4 or Stable Diffusion, are rapidly transforming the way we communicate, illustrate, and create. However, AI regulation, in the EU and beyond, has primarily focused on conventional AI models, not LGAIMs. This paper will situate these new generative models in the current debate on trustworthy AI regulation, and ask how the law can be tailored to their capabilities. After laying technical foundations, the legal part of the paper proceeds in four steps, covering (1) direct regulation, (2) data protection, (3) content moderation, and (4) policy proposals. It suggests a novel terminology to capture the AI value chain in LGAIM settings by differentiating between LGAIM developers, deployers, professional and non-professional users, as well as recipients of LGAIM output. We tailor regulatory duties to these different actors along the value chain and suggest strategies to ensure that LGAIMs are trustworthy and deployed for the benefit of society at large. Rules in the AI Act and other direct regulation must match the specificities of pre-trained models. The paper argues for three layers of obligations concerning LGAIMs (minimum standards for all LGAIMs; high-risk obligations for high-risk use cases; collaborations along the AI value chain). In general, regulation should focus on concrete high-risk applications, and not the pre-trained model itself, and should include (i) obligations regarding transparency and (ii) risk management. Non-discrimination provisions (iii) may, however, apply to LGAIM developers. Lastly, (iv) the core of the DSA's content moderation rules should be expanded to cover LGAIMs. This includes notice and action mechanisms, and trusted flaggers.